

High stakes for data-centre investors and operators: GPUaaS specialists offer growth with risk

April 2025

Jay Lee

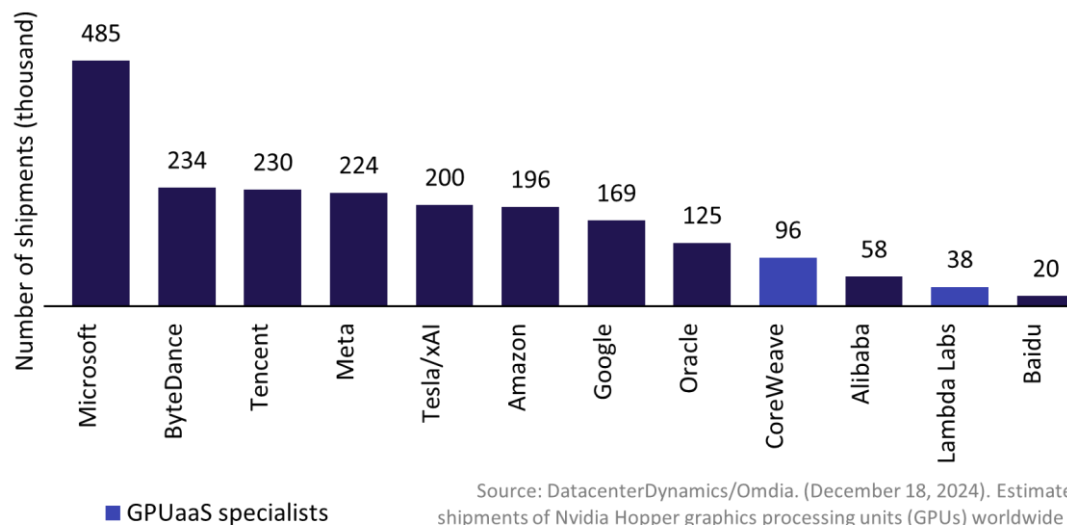
Artificial intelligence's (AI's) transformational impact on data-centre demand has spawned innovative GPU-as-a-service (GPUaaS) specialists such as CoreWeave and Lambda Labs. For data-centre operators, these tenants offer growth beyond traditional hyperscalers – but their fragmented, fast-evolving market brings risks. To capitalise, data-centre providers must adapt their infrastructure and contracts, balancing opportunity with stability.

GPUaaS specialists fuel data-centre growth

The AI boom has opened up opportunities for all sorts of new business models. Customers wanting to develop AI models need access to advanced GPUs, which can be leased from cloud service providers. GPUaaS products are now widely offered by traditional cloud service providers, but so high is the demand for access to high-power compute capacity that several GPUaaS specialists¹ have emerged. The largest players – notably CoreWeave and Lambda Labs – have raised significant funding to capture this growth.

The rapid rise of GPUaaS specialists can be seen by their share of Nvidia Hopper GPU shipments, with CoreWeave and Lambda among the 12 largest customers in 2024 and placing them in the same bracket as traditional hyperscalers (see Figure 1). In addition, while Microsoft was the largest buyer of GPUs, it was also a major GPUaaS customer – accounting for over 60% of [CoreWeave's total revenue in 2024](#).

¹ GPUaaS specialists refer to cloud computing providers who focus on leasing GPUs to support AI workloads. They are distinct from traditional cloud computing providers who have a more comprehensive suite of services that can include GPU leasing.

Figure 1: Nvidia Hopper shipments in 2024

Source: DatacenterDynamics/Omdia. (December 18, 2024). Estimated shipments of Nvidia Hopper graphics processing units (GPUs) worldwide in 2024, by customer. In Statista. Retrieved March 31, 2025, from <https://www.statista.com/statistics/1560718/nvidia-hopper-gpu-shipments-by-customer/>.

This emerging market represents a significant new segment for data-centre providers, as the GPUaaS specialists rely primarily on leased data-centre capacity to grow. CoreWeave announced in February 2025 that its leasing agreements with Core Scientific had reached 590MW in contracted capacity across six sites, reflecting the scale required to support the [high power densities of AI GPUs](#). In the Asia-Pacific region, SIAM AI (a Thai-focused GPUaaS specialist) plans for 100–200MW in data-centre capacity to [support its expansion](#).

GPUaaS is beginning to upend the traditional orthodoxy on geographical location of new data-centre hubs. Historically, data-centre demand (and thus capacity) has been concentrated at major hubs such as Frankfurt, London, Amsterdam, Paris and Dublin ('FLAP-D' markets) in Europe, and Singapore in South-East Asia. However, supply constraints in some traditional hubs coupled with latency-tolerant needs for GPUaaS is diversifying demand geographically. Markets with favourable characteristics (notably power availability and pricing) are emerging as candidates to support GPUaaS requirements – providing expansion opportunities for data-centre providers and investors. Examples of this can already be seen in developments such as [CoreWeave's 2024 announcement of plans](#) to expand into data centres in Norway, Sweden and Spain.

GPUaaS specialists amplify risks for data-centre providers

GPUaaS specialists fuel growth but can amplify risks for data-centre providers. Understanding and mitigating these risks is important for data-centre providers and their investors to consider.

Fragmented market

The GPUaaS market is still in its infancy, which breeds fragmentation – Nvidia's Partner Network [lists 40 AI-competent cloud partners](#) as of February 2025, varying by size, customer base and financial backing. Data-centre providers face counterparty risk if serving less established players whose financial runway may be contingent on them securing new major customers and/or more funding.

Technology obsolescence

GPU technology evolves quickly with constant research and development in place. New GPUs are introduced every 1–2 years: Nvidia’s Hopper architecture (March 2022) gave way to Blackwell (March 2024), with Rubin already revealed as next major release. New GPUs can shift end-user demand, forcing GPUaaS specialists to spend heavily, risking cannibalisation of existing assets and straining their ability to pay data-centre leases. The evolution of next-generation GPUs – with changing power density, weight and cooling requirements – also has a bearing on the design of data centres, which need to evolve to support these changes or risk facility obsolescence. Long-term commitments from GPUaaS customers can mitigate this risk (for example, CoreWeave’s 12-year contract with Core Scientific), but we have seen less established GPUaaS providers agreeing shorter contracts (sometimes less than 5 years in tenure) that would elevate this risk.

Exposure to GPUaaS price declines

The GPUaaS market is dynamic and has seen a significant decline in prices: by November 2024 [prices had reached](#) USD2/hour versus USD8/hour earlier in the year. With the threat of technology obsolescence looming (as described above), downward pricing pressures are likely to persist. This threatens the cash generation of GPUaaS specialists and their ability to pay for data-centre leases.

Regulatory hurdles

AI has emerged as a significant area of regulatory interest, especially in light of US-China tensions. Various measures have been introduced by the USA to limit the export of higher-spec GPUs to China and China-linked entities. In January 2025, the Biden administration published the Framework for Artificial Intelligence Diffusion, which categorises countries into three tiers – with Tier 2 and 3 [countries facing limitations on access to GPUs](#). Such restrictions may hobble or distort growth among GPUaaS specialists and thus limit their growth in co-location demand.

Mitigating GPUaaS risks is critical

Data-centre providers are faced with a significant opportunity given the substantial demand potential that GPUaaS specialists represent, but the risks require mitigation. Measures and strategies can be used to control some of these risks:

- securing a higher security deposit versus traditional hyperscale customers can guard against potential defaults
- using GPUs as potential collateral for overdue payments helps, though rapid depreciation may limit value
- ensuring revenue diversity – capping any single GPUaaS tenant’s share – reduces the risk of overexposure
- avoiding building solely for GPUaaS specialists as anchors, instead opting for adaptable facilities (for example, sites with high power density near major power and connectivity hubs) that suit hyperscalers or GPU clusters.

Controlling risks is the key to capitalising on GPUaaS’s potential

GPUaaS specialists are creating new opportunities in the data-centre market but are also driving a change in data-centre economics, demanding agility from providers and scrutiny from investors. Pursuing GPUaaS growth is key, with mitigation measures used to protect against unique risks. For data-centre investors, GPUaaS can boost growth and valuations, though careful due diligence must be made to understand GPUaaS exposure and implications. As GPUaaS matures, data-centre providers that master power efficiency and hybrid tenancy profiles that capitalise on demand from both traditional hyperscalers and GPUaaS specialists will excel.

Analysys Mason has extensive experience of technical operation and commercial models used by data-centre operators gained from projects across the global data-centre sector. We have undertaken detailed market studies, due diligence exercises (including 130+ commercial and technical due diligence projects on data-centre and cloud assets since 2019) as well as strategic support to data-centre operators and their investors. For further information, please contact Jay Lee.